Characterizing the Spatial Structure(s) of Cities "on the fly": the Space-Time Calendar

Dani Arribas-Bel[a*] Emmanouil Tranos[b†]

[a] *Department of Geography and Planning, University of Liverpool, United Kingdom*
[b] *School of Geography, Earth and Environmental Sciences, University of Birmingham, United Kingdom*

July 20, 2017

Abstract

Our understanding of the spatial structure of cities has been traditionally shaped by the availability of static data. In the last few years, thanks to improvements in geospatial technology as well as computing storage and power, there has been an explosion of geo-referenced data, which monitor cities and urban activities in real time. Although this shift in the data landscape promises to change and augment the way we measure, understand, and act on cities, it poses significant methodological challenges and uncovers substantial gaps in the analytics required to leverage its power. The present paper contributes to this agenda by delivering insights in two fundamental fronts: first, we compare several methods that conceptualize both space and time in rather different ways, highlighting their main advantages and limitations; second and more important, we propose a novel approach –the Space-Time Calendar– that uncovers, characterises and visualizes in an explicitly spatial way both *fast* and *slow* urban dynamics. We illustrate the advantages of the Space-Time Calendar using a dataset derived from over two years of mobile phone activity in the city of Amsterdam (The Netherlands). Our findings highlight the advantages of the Space-Time Calendar approach, but also the benefits of appropriately matching the methodological approach to the nature of the data at hand.

Introduction

Data availability has historically been a concern for urban studies. Traditional sources were usually limited to a finite number of official outlets such as censuses, travel surveys and administrative data. Given the nature and characteristics of available data, urban researchers have developed a plethora of robust methodological approaches to help themselves in answering questions related to cities and the processes that take place within their boundaries. Consequently, such techniques are tuned to extract as much knowledge as possible from the shape and features of the available data. Given the available technology, datasets used in urban research have traditionally been constrained by a substantial degree of aggregation over space (large areas) and/or time (with very low frequency) as well as by limited coverage (small population samples). The development of such methodologies can be traced back to the

*D.Arribas-Bel@liverpool.ac.uk
†E.Tranos@bham.ac.uk

founding of quantitative geography (e.g. Von Thunen, 1826; Christaller, 1966; Hagerstrand et al., 1968), a stream of research which has been identified by Portugali as the *first culture of cities* [1] (Portugali, 2011).

The XXIst. Century has brought a series of technological advances that are reshaping the urban data landscape. As documented in Arribas-Bel (2014), the combination of affordable computing power, ubiquitous connectivity, and cheap geospatial technology embedded in mobile devices has created unprecedented amounts of data about urban life. These *digital bread-crumbs* (Rabari and Storper, 2014) of the digital revolution conform to very different characteristics than the data urban research is used to but, nevertheless, pose a unique opportunity for the part of the social sciences interested in cities (Batty, 2013a). In fact, the emergence of such new forms of data has ignited a heated discussion about their ontology, value and limitations within the urban domain (see for example Bettencourt, 2014; Batty et al., 2012; Kitchin, 2013; Townsend, 2013).

Despite numerous epistemological debates and a growing set of empirical applications of new forms of data (see for example Calabrese et al., 2014; Blondel et al., 2015; Steenbruggen et al., 2015 as well as Kitchin, 2014b and Kitchin, 2014a), the methodological toolkit employed by researchers is still heavily influenced by the data attributes of the first culture of cities. Although the emergence of diverse sources of big data may have the potential to enable urban researchers to ask new questions or answer questions that traditional data sources did not enable them to do, it is not clear that approaching them with standard techniques gives rise to the best possible results. Methodologies originally developed to analyse *small data* are not necessarily equipped to tackle some of the distinctive features of newer sources. As Kitchin puts it, "[i]t is thus clear that further research is required to adapt, hone and extend existing techniques and to invent new methods that can make sense of and extract value from big data and data infrastructures" (Kitchin, 2014b, p.112). Before identifying areas where traditional analytical methods may not be adequately equipped to fully exploit new forms of data, it is worth discussing their main characteristics. According to the classification in Arribas-Bel (2014), three broad categories of big data can be identified: (i) data derived from individuals with mobile devices (the "citizens-as-sensors" paradigm, Goodchild, 2007); (ii) data from businesses online activities; and (iii) government open datasets. Despite their diverse nature, Kitchin (2013) identified a number of common attributes based on a survey of the relevant literature (boyd and Crawford, 2012; Dodge and Kitchin, 2005; Laney, 2001; Marz and Warren, 2015; Mayer-Schonberger and Cukier, 2013; Zikopoulos et al., 2012): massive volume, high velocity, diverse variety, exhaustive scope, high resolution, relational in nature and flexibility. The above contrasts significantly with *traditional* sources of aggregated, infrequent, reductionist and inflexible urban data that researchers of the last century were usually limited to.

The availability of traditional sources of data has spurred much research on what has come to be known as the *slow dynamics* of cities: the kind of evolution that can only be observed over longer periods of time. The rise of new forms of data is shifting some of the emphasis towards *fast dynamics*: patterns, trends and changes that take place within short spans of time and relate to issues around mobility and flows (Batty, 2009, 2013b; Wegener et al., 1983; Snickars et al., 1982). In its extreme, this can lead to what Batty (2013b) calls "short termism", if the focus on the immediate relegates the study of longer term effects to a second level of attention. But, with awareness of this risk, the potential of this approach in obtaining a much more complete understanding of how cities function is clear. However, "short-termism" is not the only issue raised by the focus on fast dynamics; methods have also been a limiting factor. The lack of temporally fine-grained data about cities and the focus on slow dynamics has made researchers pay much less attention to the temporal dimension of techniques to measure the spatial structure of cities. As An et al. (2015) put it, urban research tends to approach space and time disjointedly, regardless of the consensus on the unified nature of space and time in many areas of geographical research. This methodological gap should not come as a surprise since

[1] While the first culture of cities refers to urban research based on hard scientific approaches, the *second culture of cities* recognises the humanistic turn in urban studies influenced by David Harvey's ideas, and the *third culture of cities* bridges the gap of the above two by employing the Complexity Theory of Cities (Portugali, 2011; Snow, 1964)

the “first culture of cities” type of urban methods were developed to tackle data of low spatial and temporal resolution. Consequently, some of the most broadly used quantitative methods only speak at the temporal dimension in an ex-post manner by superimposing cross-sections of data at different time intervals. This is not only a characteristic of urban research, but analogous concerns have been raised in other areas of geographical research including health, segregation, and accessibility studies (Kwan, 2013). The advent of the urban data revolution and the inclusion of fast dynamics in the research agenda make these gaps all the more evident.

Our paper addresses several of these methodological gaps in a timely manner, delivering two key contributions. First, we compare and assess the ability of widely used methods to extract useful insights on urban dynamics from new forms of (space-time) data. Secondly, we propose a novel methodological approach to capture both fast and slow dynamics, enabling the user to take full advantage of the spatial and temporal granularity afforded by such data in an explicitly integrated fashion. As Richardson (2013) puts it, the availability of big data sources is only going to be increased in the not so distant future and therefore urban researchers need to adapt older and develop new methods capable of tackling the (spatio-temporal) granularity of such data. We focus on urban spatial structure given its relevance and wide range of applications, and present a new approach that extracts insights from large amounts of new data. In the process, we also assess how well different conceptualizations of space and time can shape and influence the analysis of spatial dynamics in cities. Emphasis is placed first on characterizing fast dynamics. In particular, we consider the added value of explicitly incorporating time in the analysis, instead of ignoring it or including it in an ad-hoc manner. This comparison is carried out against the background of different conceptualizations of space. We then “zoom out” and switch our focus of attention to slow dynamics. In this context, our interest is to understand the evolution of fast processes over longer periods of time. With the goal in mind of connecting these two temporal scales, this paper proposes a new visualization device we term the Space-Time Calendar. We show how the Space-Time Calendar can uncover patterns and insights that would otherwise be very hard to elucidate by applying it to the case of Amsterdam’s spatial structure. We also note the Space-Time Calendar itself, however, is a flexible tool whose range of applications spans a much wider set of contexts. The remainder is structured as follows. The next section briefly reviews the literature on urban spatial structure as well as applications using new forms of data in that direction. The third section provides the methodological background, setting up the comparison of different space(-time) techniques and introducing the Space-Time Calendar as the novel contribution of this paper. In the fourth section, we present an application of the Space-Time Calendar to a large dataset of mobile phone activity in the city of Amsterdam. We conclude on the final section.

The spatial structure of cities: “old” and “new” data

The study of urban structure and its dynamics provides a fertile ground for comparing different methods developed with different types of data in mind. Traditionally, research on the spatial structure of cities has relied heavily on cross-sectional data. Most discussions about its changes or evolution over time has been based on infrequent data, such as decadal censuses. The above characterises a whole generation of studies starting from the seminal work of Von Thunen (1826), the Alonso-Muth-Mills model (Alonso, 1964; Mills, 1972; Muth, 1969), but also more recent studies such as McDonald (1987); Gordon and Richardson (1996); Anas et al. (1998), and McMillen (2001). Summarising the empirical literature on urban structure, Arribas-Bel and Schmidt (2013) highlights some of its key dimensions, which include, among others, population density, employment concentration, size, polycentricity, land-use mixing and commuting flows.

The methodological approaches adopted in these studies are geared towards capturing the *slow dynamics* of cities or, in other words, the changes taking place over longer periods of time, as reflected in low-frequency data. The development of spatial statistics and econometrics supported this stream of research and provided flexibility in order to recognise urban centres without assumptions and predefined knowledge about the study area and its structure (Páez and Scott, 2005; Baumont et al., 2004; Pereira et al., 2013). Such methodologies enhance the identification of hot-spots in the form of statistically significant spatial clusters. For example, Paez et al. (2001) studies land price variation using the $G_i(d)$ family of statistics, and Baumont et al. (2004) employs local indicators of spatial association (LISA, Anselin, 1995) to identify population and employment concentrations. LISA and other exploratory spatial data analysis (ESDA) type of methods have been widely used in such research. Recently Arribas-Bel and Sanz-Gracia (2014) investigated employment centres in the US over the course of thirty years and Salvati et al. (2016) employed ESDA to study changes in built-up areas. In addition, similar methods have been applied in studies related to employment subcentres (Guillain et al., 2006), property values in polycentric cities (Han, 2005), and the structure of urban land uses (Salvati and Carlucci, 2014), to name just a few.

A key characteristic of the above studies is the use of aggregated variables represented usually as spatial polygons (e.g. population in census tracts). LISA and related techniques accommodate well such data as it was originally conceived to be applied on lattice data. Common as this type of data might be, urban processes are also represented by other types of spatial data. Kulldorff's Scan statistic (Kulldorff, 1997) and its subsequent space-time extensions (e.g. Kulldorff et al., 2005, 2009) were originally designed for epidemiology studies based on point data. This approach is capable of identifying significant space(-time) clusters in the shape of circles, ellipses, and cylinders. Despite being originally developed for applications in a different field, the increasing availability of space-time data and the urge to understand the dynamics of cities has increased the popularity of this method among urban researchers. For instance, Cheng and Wicks (2014) applied Space-Time Scan Statistics (STSS) on a Twitter dataset to identify space-time significant clusters which represent references to specific events. Kang (2010) employed STSS as a means to prove that space-time clustering better represents agglomeration economies and their evolution than spatial clustering. In a similar vein, Tuia et al. (2009) used STSS to identify significant space-time clusters in the distribution of the residential patterns of professions in Switzerland. More methodological papers on the use of STSS in urban research include the work of López et al. (2015), which evaluated the effectiveness of STSS in specification testing for spatial econometric models and more specifically for hedonic models, as well as the work of Cheng and Adepeju (2014) who expanded the notion of the Modifiable Areal Unit Problem (MAUP) by incorporating time using STSS. All the above studies are situated within the broader developments that the space-time analysis research field has achieved in its endeavour to analyse and model space-time data. These developments, which are systematically illustrated by An et al. (2015) include various methods which are used to (i) identify patterns, (ii) develop space-time statistical models and (iii) simulate processes for either individual movement data or spatial panel data (see Table 2 in An et al., 2015). Our paper contributes in the *pattern revelation* for *spatial panel data analysis* to follow An et al. (2015) typology. Recent developments within this category include, among others, the work of Rey and Janikas (2006), Lee et al. (2014), Delmelle et al. (2014) and Ye and Carroll (2011).

Part of the reason why new methodological approaches have been sought is that new forms of data required slightly different perspectives. Examples of such sources which are used in order to understand cities and their spatial structure include: location-based services (e.g. Ferrari et al., 2011; Arribas-Bel et al., 2015), smart transit cards (Roth et al., 2011), real estate online listings (Rae, 2015; Rae and Sener, 2016; Boeing and Waddell, 2016), or social media (e.g. Arribas-Bel, 2015), to name just a few. Given their pervasiveness and high degree of penetration among urban populations, a substantial amount of empirical research has utilised data from mobile phone operators with this purpose. Reades et al. (2009) detected a strong relationship between human activity in cities and aggregated mobile phone usage. Louail et al. (2014) used similar data to analyse the structure of 31 Spanish cities and, earlier, Ratti et al. (2006) had mapped urban activities in Amsterdam and Rome using mobile phone usage. In the same vein, Sevtsuk and Ratti (2010) estimated the distribution of urban population over space and time. Using aggregated data on mobile phone usage, Jacobs-Crisioni et al. (2014) assessed the effect of land-use mixing on urban activity patterns, while Tranos and Nijkamp (2015) modelled the space-time dynamics of aggregated human activity in Amsterdam. Several other researchers have used similar sources as a way to determine land use (e.g. Reades et al., 2007; Toole et al., 2012; Pei et al., 2014) and human mobility within cities (Blondel et al., 2008; Licoppe et al., 2008; Calabrese et al., 2013).

Methodological framework

Conceptualizations of space and time

Our firt interest is finding out whether, given a dataset of much higher resolution than what used to be common, it is also necessary to "upgrade" the methodological approach to extract new insights, or the data are able to "speak by themselves" even with traditional techniques. In this context, higher resolution operates along two main dimensions: space and time. Over space, the increase in granularity means that, for a given area of study, it is now possible to obtain many more observations, dividing space into ever smaller portions. Over time, new datasets increasingly include a temporal dimension that allows to place measurements not only at locations but at given moments in time. Additionally, such timestamps can be equally fine grained, offering the possibility of reconstructing a detailed quantitative description of how urban phenomena unfold. In both cases, each dimension can be understood as either a succession of discrete partitions or as a continuous trend. Traditionally, urban research has relied on the former because the coarseness of data made it unrealistic to assume the latter. However, it is not unreasonable to think that, with the levels of detail afforded by recent datasets, a continuous conceptualization might prove beneficial. We focus our strategy on two main aspects, represented along the vertical and horizonat axes of Table 1: to what extent the degree of sophistication in the way the temporal dimension of the data influences the depth of insights obtained from the analysis; and in which ways, if any, the formalization of space into the statistical method used affects the outcomes too.

	Discrete	Continuous
Space only	LISA	Spatial Scan
"Pseudo" space-time	Repeated LISA	Repeated Spatial Scan
Space-time	Space-Time LISA	Space-Time Scan

Table 1: Methods comparison

Our exercise begins by creating a baseline that ignores time completely. The rationale behind it is twofold: on the one hand, it provides an initial benchmark that makes it easier to assess the value added of introducing the temporal dimension; on the other hand, it serves as an illustration of the most common approach the literature has adopted to consider the spatial structure of a city, usually constrained to a purely spatial approach by the availability of data. For the discrete understanding of space, we rely on the widely used LISA statistics, proposed by Anselin (1995), which can be expressed as:

$$I_i = \left(\frac{z_i}{m_2}\right) \sum_j w_{ij} z_j \tag{1}$$

where z_i is the standardized value for observation i (raw measurement minus the average in the dataset and divided by its standard deviation), m_2 is the second moment (variance) of the variable and w_{ij} is the ij-th cell of the spatial weights matrix W that captures whether observations i and j are spatial neighbors ($w_{ij} > 0$) or not ($w_{ij} = 0$). In our case, following common practice, we use a queen contiguity criterion by which i and j are neighbors and initially receive a weight of one ($w_{ij} = 1$) if their boundaries share at least a point. Once built, we standardize W so every row sums to one, effectively making the second component of Equation 1, the spatial lag of i, the average value of z in i's neighborhood. Inference on I_i can be performed through normal approximation but, more often, is calculated using a permutation approach (e.g. Anselin et al., 2006) or through bootstrap (a Yan et al., 2015). One of the strengths of the LISA is that it allows to identify not only pockets of positive spatial autocorrelation (clusters of either high or low values), but also "spatial outliers": areas with low values surrounded by high values, and vice versa.

Conceptualizing space into W was originally created for lattice data, a case where the observations are given, fixed, and have a clear structure already. The approach, however, forces the researcher to further impose an exogenous spatial configuration. Decisions about who is neighbour with who and "to what extent" they are neighbours (i.e. is $w_{ij} > 0$ and, if so, what its value is) need to be made prior to any analysis and, to some extent, influence its final outcome. On the other hand, using a W to formalize space makes such assumptions explicit and, once such structure is created, the LISA is flexible enough to produce irregularly-shaped clusters. This is particularly the case if the number of observations considered in the analysis is large and hence their size small: in this case, a simple W connecting each observation to its immediate neighbours provides a large degree of flexibility when it comes to delineating the boundary of a cluster.

The continuous counterpart of the LISA statistic above is represented by the purely spatial Scan for Gaussian processes, as suggested by Kulldorff et al. (2009), which is a variant of the original Scan statistic presented in Kulldorff (1997):

$$LLR(r) = max_r \left(\frac{\ln L_r}{\ln L_0} \right) \tag{2}$$

where r represents a circle centred around a given location, L_0 is the likelihood under the null hypothesis of mean stability across the dataset and L_r is the likelihood calculated under the alternative, which assumes there are spatial clusters with values higher than in other locations. Respectively, these can be written as:

$$\ln L_0 = -N \ln(\sqrt{2\pi}) - N \ln(\sigma) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$\ln L_r = -N \ln(\sqrt{2\pi}) - N \ln(\sqrt{\sigma_r^2}) - \frac{N}{2}$$

where N is the total number of observations, and μ and σ are the maximum likelihood estimates of the mean and the variance, respectively. The term σ_r^2 refers to the variance under the alternative. The radius and shape of circle r is selected among a large number of candidates as that which maximizes LLR. We follow the original suggestion to centre all the circles in the observations in the dataset. Once a candidate circle is identified, inference to assess its probability of arising from a random process is created through a permutation approach that randomly shuffles values x_i and their location creating, similar to the LISA approach, a pseudo p-value. The methodology allows identification of clusters of either high or low values.

The Scan approach was originally designed to identify clusters of events that can occur anywhere in space. Indeed, the original proposal by Kulldorff (1997) was designed for point processes. As such, it does not rely on locations that are given and fixed over space, but rather it is based on drawing boundaries within the full area extent and analysing values within and outside the window. This continuous treatment of space is in a way much more attractive in principle as it does not require the researcher to specify any particular structure prior to the analysis. Additionally, it is also more data-driven in that the window selection is largely based on the maximization of a likelihood ratio calculated based on the values of the dataset. However, windows need to conform to the shape of a circle (or ellipsoid in some cases), and this imposes a spatial structure that may preclude the identification of clusters with highly irregular shapes or, when they are found, their true shape may be obscured by that of the circle used to identify it.

Moving on to introduce time, our first stage is what we term "pseudo space-time". In this case, the time dimension is recognized, albeit not included in the computation explicitly, but rather in an ah-hoc manner. In the LISA case, we apply Equation 1 to values for each given hour in the day. Similarly, we apply Equation 2 to the cross-section of values for each hour to obtain hourly clusters using the Scan approach. It is important to note that in both cases, this is a "pseudo" space-time approach as, at the moment of the computation of the clusters for a given hour, only variation within that time interval is considered.

To introduce time as a "first-class" citizen in the discrete approach, we use Space-Time extension of Anselin (1995), as presented in Lee and Li (2016):

$$I_{it} = \left(\frac{z_{it}}{m_2}\right) \sum_j w_{it-ju} z_{ju} \tag{3}$$

where everything stays as in Equation 1 except, in this case, the entire panel of location-hour data is pooled for the calculation of the clusters for each hour. Consequently, the it suffix represents observation i at time t, within the context of the entire space-time dataset. Equally, W becomes an explicit space-time adjacency matrix that captures neighborhood relations both over space (e.g. sharing borders) and time (e.g. previous/following period). In that context, j represents a spatial unit different than i while u represents a time period different from t. Our choice for W in our analysis is to stack the cross-sectional matrices for each hour into a full day, space-time weights matrix, filling the remainder of the matrix with zeros.

In the continuous case, the space-time generalization of Equation 2 is much closer in notation to its cross-section counterpart:

$$LLR(c) = max_c \left(\frac{lnL_c}{lnL_0}\right) \tag{4}$$

where the only change is that the definition of a cluster is now delimited using a space-time cylinder that may include observations within a close spatial (e.g. metres) or temporal (e.g. hours) distance.

Although not the main focus of this paper, the integration of space and time deserves further comment. The transition from purely spatial analysis, as presented first in this section, into truly space-time considerations demands the reconciliation of the "dual nature" (An et al., 2015) of the two fundamentally different dimensions, expressed in different scales. The two ends of the spectrum considered above –discrete and continous– to conceptualize space very much apply in this context. On the one hand, the LISA approach allows and demands the division of time into ex-ante units (e.g. hours, days, years) and, similarly as with spatial units, requires an exogenous categorization of (temporal) neighborhood relationships. This strategy can be desired in cases where either the data or theoretical considerations suggest an ex-ante definition of such units and relations. On the other hand, the scan-based approaches offer a more data-driven strategy that follows rather naturally from their continuous approach to space: by extending spatial circles into space-time cylinders, defining several of them, and selecting that which maximizes the likelihood ratio in Equation 4.

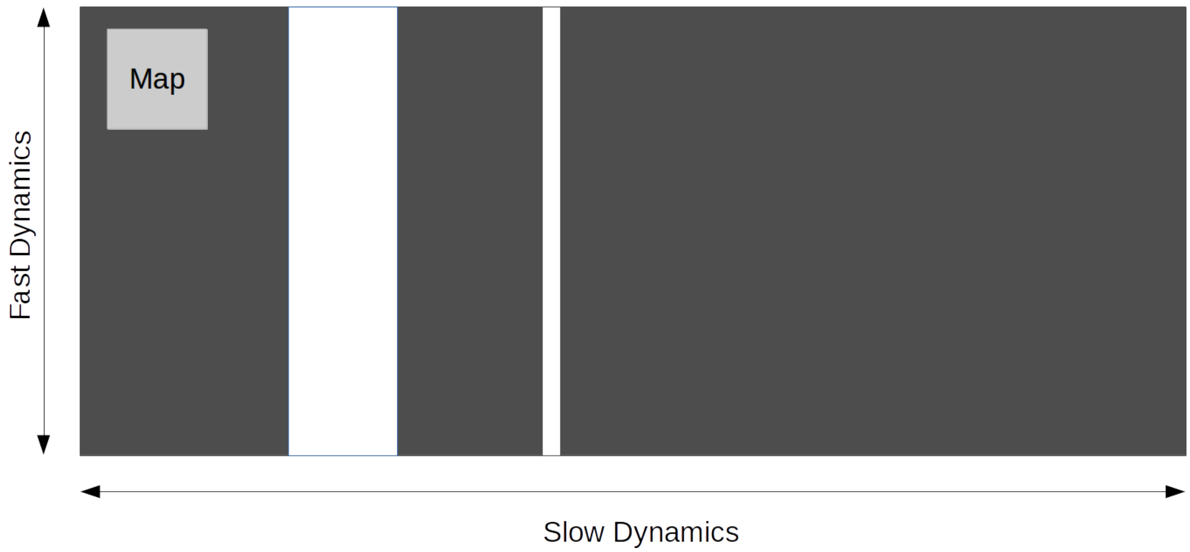


Figure 1: Space-Time Calendar

The Space-Time Calendar

The approaches outlined above afford a greater degree of detail when looking at the internal dynamics of a city. They allow to expand the view to focus on changes that occur within a single day, hence revealing patterns that were otherwise hidden by time aggregation. However, the space-time LISA's (and that of similar methods such as the STSS) very best property – temporal resolution – becomes its most salient disadvantage if the data spans over a long period of time. The insights the technique allows come at the cost of the display of a large amount of information. The approach is effective if the data is aggregated, for instance, into an average day. But as soon as the analysis aims to include more than a single set of 24 hours, the traditional visual strategy displaying a map for each hour becomes inefficient. Their overall stability makes it hard to gauge differences and anomalies across a large number of mostly similar figures with 24 maps each. Moreover, the processes which result to fast changes in the shape of urban structure coexist with processes which underpin slower urban dynamics. Therefore, reducing the temporal variability of years' worth of hourly urban data to 24 maps representing the average day masks away the processes which result to slower changes of urban structure. Maps for every single hour over several days, weeks or months should thus be replaced by a more effective visual strategy. An ideal approach should be able to pick up the daily rhythm and the fast dynamics of cities and, at the same time, it should also highlight explicitly the changes that occur in such pattern over longer periods of time.

With that purpose in mind, we introduce the Space-Time Calendar, a flexible graphical device to visualize the analysis of large volumes of spatio-temporal data. The Space-Time Calendar organizes information about a single area produced from several runs of a space-time statistic such as the space-time LISA or the STSS. The structure of a Space-Time Calendar is generically displayed in Figure 1, and is based on a modified rectangular heatmap where the vertical axis is used to represent temporal units within a single run of the space-time statistic, while the horizontal axis stacks several runs over a longer period of time. This layout graphically encodes the two main conceptualizations of urban dynamics reviewed above: fast and slow. One can thus read a Space-Time Calendar vertically to characterize slow dynamics and, once a profile has been drawn up for a given area on their basis, one can move along the horizontal axis to analyse its evolution over longer periods of time, effectively obtaining a representation of the slow dynamics of the system. In order to make the translation from taditional LISA/STSS maps to the Space-Time Calendar easier, a similar colour scheme is used to fill in each cell corresponding to a given time unit for a given run, leaving it blank if no data is available for that particular point in time (dark grey and white, respectively, in Figure 1). Finally, a small simplified map of the full geographical extent considered in the analysis, highlighting the area the Space-Time Calendar depicts, is included in light grey in the upper-left corner. Because the approach is based on an explicitly spatio-temporal technique, a single Space-Time Calendar captures information not only about the particular area it represents, but also expanding into its immediate geographical and temporal context, as defined by the space-time weights matrix/cylinder used. Moreover, it is also an effective tool to quickly spot “holes” in the data as the absence of results at a given day evidences them.

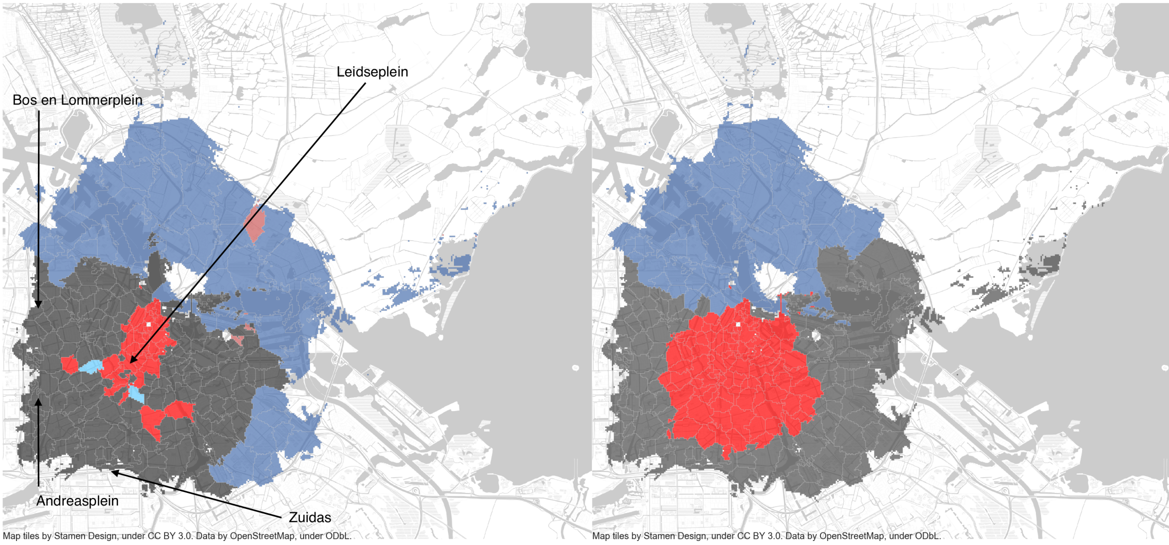
In the context of our application, the Space-Time Calendar provides an intuitive way to visualize large amounts of daily data on urban activity. Specifically, it proposes a solution to the overload of information that results from trying to scale up the traditional approach, thus allowing to extend it to more than a single day of data. In essence, every cell of the Space-Time Calendar in our example will display the pairing of a given hour along the vertical axis (*fast* dynamics) and a given day along the horizontal one (*slow* dynamics), combined with a colour scheme that represents whether the area, at that moment, was part of a cluster of high/low activity, a spatial outlier, or a non-significant observation. It is important to note that the horizontal axis is expressed in days and spans over the entire coverage of the data. This means that, when data are not available, the Space-Time Calendar leaves a blank space. Compared to the gradient of a traditional heatmap, the colour coding applied helps to simplify the view and focus on the key elements relevant to the function of an area at a given point in time (*Is it a hotspot? A coldspot? A spatial outlier? Or is it not part of any cluster?*).

Empirical application

This section introduces the data employed for this paper as well as the results of the analysis. The dataset was provided by a major mobile phone operator in The Netherlands and contains aggregated telecommunication counts at the level of the GSM (Global System for Mobile communications) cell. These zones represent the polygons each GSM antenna serves. They are irregular zones with varying size, in a way that smaller GSM zones are designed for busier areas. These data are made available to us at an hourly basis for the period of December 2007 – November 2010, although there are a few months for which we do not have data and some random days that display a few missing measurements (see also Jacobs-Crisioni et al., 2014; Tranos and Nijkamp, 2015; Steenbruggen et al., 2016 for a more detailed description of the dataset). In total, 223 such urban zones are included in the analysis and all of these zones are contained within A10, Amsterdam’s main ring road (see Figure 2 for general reference). The main telecommunication count that the analysis focuses upon is the number of Erlangs. This is an aggregate unit of telecommunication activity which is equivalent to 60 minutes of a voice phone call: for example if 20 phone calls take place within one hour at a specific GSM cell and each of these calls has a duration of 6 min, then the total number of Erlangs will be 2 ($20 * 6 = 120$ min = 2h). Given the large variation in the size of each zone, we use Erlang density as the main variable of focus.

Taken altogether, the assembled dataset provides an exceptional benchmark to compare space-time methods in the context of urban research. Additionally, the nature of the dataset resembles that of many new forms of data increasingly becoming available about urban activity: rich in detail both over time and space, although fixed over space at a given set of locations (e.g. related to fixed location of sensors), thus available in different geographical resolution than it is common for official sources such as censuses, and originally created with a purpose in mind other than research (i.e. "accidental", Arribas-Bel, 2014).

Figure 2 presents the comparison between the LISA and the Scan method, when all the temporal variation is compressed into a single cross-section. The results derived from the two methods are substantially different, albeit representative of each method’s core attributes. The Scan method (Figure 2 b), built around a continuous understanding of space, generates clusters with the shape of the circles



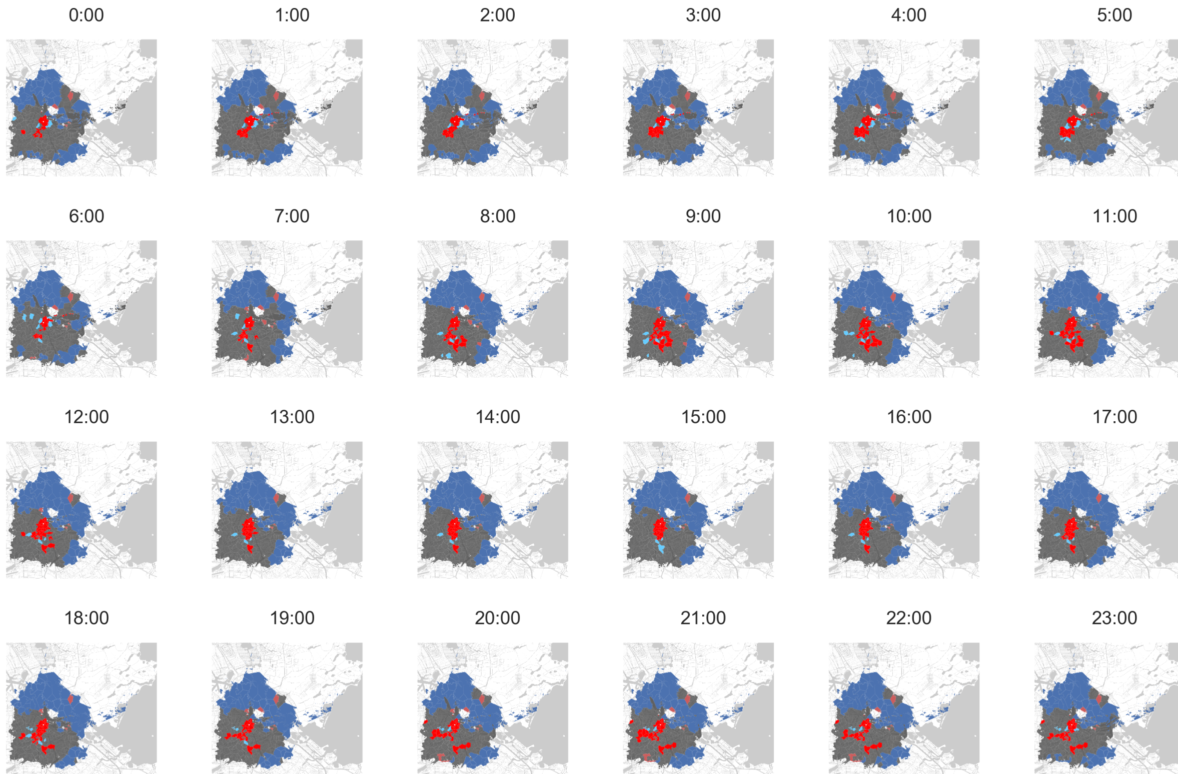
(a) LISA (b) Scan

Colour scheme: dark red for High-High significant clusters; dark blue for Low-Low significant clusters; light red for High-Low significant spatial outliers; light blue for Low-High significant spatial outliers; dark grey for non-significant.

Figure 2: Cross-section results

used to search of concentration of values. However, the data used for this exercise are discrete and aggregated, with their areal nature representing GSM zones within the city of Amsterdam. As the figure illustrates, the mismatch between method and data gives rise to a rather coarse delimitation of Amsterdam's main cluster of activity. In order for the statistic to include the main hotspots (red), it needs to enlarge the circle in a way that encompasses most of the southern part of the city, making it less discriminative and, in that sense, useful. Similarly, some potential areas in the mostly residential northern part of the city are not identified as a coldspot because the circle would need to be so large that parts of the super active city centre fall within its boundaries. On the contrary, the LISA accommodates discrete data much better due to the flexible delineation of space encoded in the spatial weights matrix, once this is created. As Figure 2 (a) illustrates, this attribute translates into a much more detailed definition of clusters of high mobile phone usage density. In other words, because the LISA's assumptions and features in understanding space match the nature of the dataset employed, its results are more satisfactory than those from its continuous counter-part, the Scan.

The above is not criticism against the Scan method, but rather a reflection of the value of selecting an appropriate technique that "matches" the nature and characteristics of a given dataset. The Scan method is tuned towards identifying clusters of events which can occur anywhere in space but, given the location of the observations in this case if fixed, this does not translate in tangible benefits. What it means however is that, the "price" the method pays to be able to accommodate locations that are not pre-specified –the circles used– becomes "expensive" in this context, as it forces the clusters into a particular shape that may not fit the underlying structure of a given cluster. Therefore the Scan method is not an optimal approach for our data. In contrast, the nature of the data we use and the scale of our analysis lead us to the adoption of the LISA type of statistics as they are deemed more appropriate for our analysis. As an additional methodological bonus, the LISA allows the identification of spatial outliers, areas of significant difference in their value with respect to those of their neighbours. Although, in a static picture of the spatial structure of a city, this might not be particularly interesting, their relevance will become clearer once we introduce a dynamic element later in the section. Given the reasons argued above and in the interest of space constraints, in the remainder of the section we will



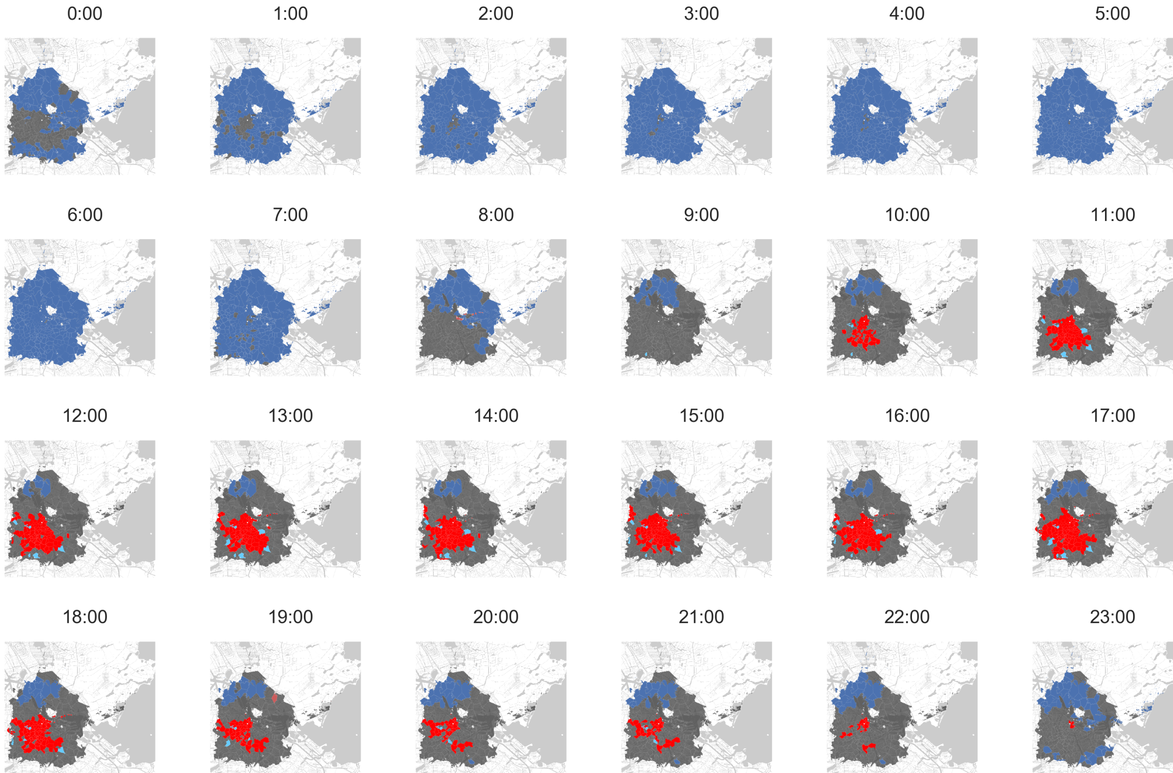
Colour scheme: dark red for High-High significant clusters; dark blue for Low-Low significant clusters; light red for High-Low significant spatial outliers; light blue for Low-High significant spatial outliers; dark grey for non-significant.

Figure 3: LISA repeated cross-sections

only show results based on the LISA type of approach. Similar results for the Scan are available from the authors.

Figure 3 presents the LISA-based pseudo space-time analysis. The temporal dimension is indirectly recognised in this case as LISA is applied on twenty-four cross-sections, each of which represents the average Erlang density of the specific hour during the overall study period. For example, the map for 0:00 in Figure 3 illustrates the significant clusters of high mobile phone usage derived from the LISA analysis on the average mobile phone usage between 0:00 - 0:59 during the study period. This set of 24 maps sheds more light on the fast dynamics of Amsterdam's spatial structure than Figure 2. Indeed, during night hours only a handful of GSM zones are part of a high cluster and the variation among the different times is minimal. From 7am onwards more GSM zones turn on as significant clusters of high activity. In other words, more polygons appear to play central roles for the city of Amsterdam at 9am than at 1am. Such a spread of disconnected high clusters within the central area characterises Amsterdam's spatial structure until 1pm. After that point we observe a higher concentration of hotspots around the geometrical centre, and this pattern evolves once the working day ends, around 6pm, into another polycentric spread of clusters of high activity, likely reflecting the importance of residential areas.

Following Table 1, the next step in addressing the role of the temporal dimension is the implementation of an explicit Space-Time LISA, which is presented in Figure 4. In this case, time is built into the core of the method as the entire panel of GSM zones and average Erlang density is pooled across the 24 hours of a representative day. Figure 4 illustrates several of the advantages of the Space-Time LISA approach. On the one hand, it identifies very well the differences between night and day patterns,



Colour scheme: dark red for High-High significant clusters; dark blue for Low-Low significant clusters; light red for High-Low significant spatial outliers; light blue for Low-High significant spatial outliers; dark grey for non-significant.

Figure 4: Space-Time LISA

which are expected and intuitive, but it is nevertheless reassuring that the algorithm picks them up so clearly. At the same time, the figure illustrates perfectly some of the advantages of a more granular dataset and begins to unpack some of the fast dynamics of the city that go completely unnoticed by considering a fully compressed view, as well as some of the benefits of including time as a “first class citizen”. Although the main hub of activity is similar in all three cases, focused around the canal rings, Figure 4 depicts more efficiently how the centre grows and shrinks over day and night. The explanation for this lies on the methodological differences between the methods. By considering the variation within the span of a whole day, the Space-Time LISA is able to unveil different rythms and cycles that the cross-section is completely unable to show and the “stitching” of cross-sections cannot pick up because the map of each hour was calculated without taking into consideration the rest of the day. Therefore, clusters of high activity we can observe in Figure 3 during night times do not appear as significant in Figure 4. In the same vein, the whole central part of Amsterdam appears to be a significant cluster of low activity during night as any mobile phone usage that takes place at night is very small in comparison to daily activity. The space-time LISA approach makes visually explicit the fact that activity is not always on, especially if we compare it with the repeated cross-section approach in Figure 3. In addition, the space-time LISA effectively illustrates how the centre evolves over the course of an average day in certain areas (central station and west of it), but fades at the end of the day in non-symmetric ways. Most notably, although the morning increase takes place mostly in a monocentric way, with the hub growing outwards, the night fading occurs in a much more polycentric fashion, with activity shrinking into a few distinct hotspots. In other words, Figure 4 demonstrates that, beyond a single spatial structure, the city has several of them that differ in important ways depending on the time of the day.

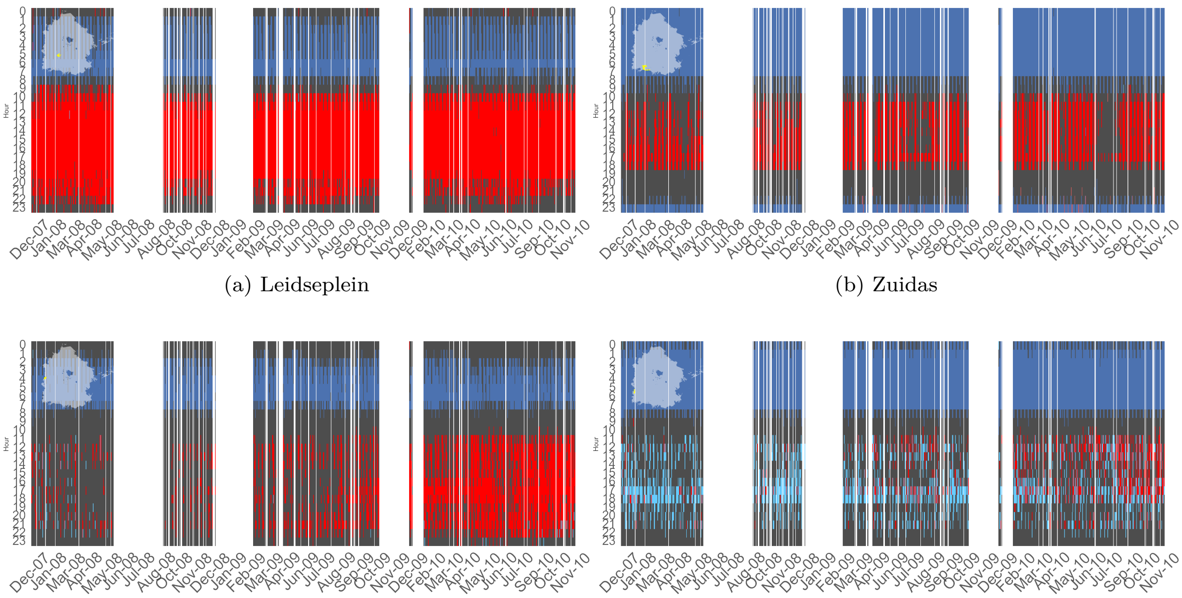
Once we have presented the results for an average day, and decided the Space-Time LISA represents the most appropriate methodological choice to explore the elements we are interested in about fast dynamics in cities, we move on the the analysis of the entire granular dataset. We do this by expanding our capacity to cope with and comprehend fast urban dynamics over longer periods of time, thus unveiling patterns which otherwise would have gone unobserved. As described above, we introduce a new graphical device –the Space-Time Calendar– that will enable us to visualize the "slow" evolution of "fast"dynamics. The Space-Time Calendar view will enable us to visualise in a concise way the fast dynamics of Amsterdam's spatial structure over the long run. This approach essentially unlocks the use of new forms of granular and extensive data like mobile phone usage to analyse urban dynamics.

Figure 5 illustrates these trends in four neighbourhoods using the Space-Time Calendar: Leidseplein (a), Zuidas (b), Bos en Lommerplein (c) and Andreaseplein (d) – see see Figure 2 for the exact locations. Leidsplein is a central area, which endures its role as a hotspot both during day and night times (9am to 10pm approx.) throughout the entire period of study. This is not surprising for anyone who has visited Amsterdam as Leidseplein is one of the most vibrant areas in central Amsterdam, both during day and night. Most importantly, its central role has not been challenged at all during the study period.

A slightly more nuanced picture can be observed for Zuidas. This is one of Amsterdam's main business and finance area, home of the World Trade Center building, and is located in the southern tip of the city. The identification of this neighbourhood as a hotspot during working hours for most of the days in our study period, as illustrated in Figure 5, demonstrates the role of this area as an employment centre. This is further enhanced by the lack of continuity in the red strips of Zuidas in the Space-Time Calendar. Zuidas is not a cluster of high activity for every day of our study period (as Leidseplein in Figure 5 b is, for example), but there is a periodic lack of significant space-time LISA clusters. Looking back on the space-time LISA estimations, we can confirm that these cycles represent working and non-working days and Zuidas only performs central functions during working days. This observation reflects work cycles and will not come as a surprise for anyone with local knowledge for Amsterdam. In addition, Figure 5 (b) enables the understand of how the mid-term dynamics of Amsterdam affect Zuidas's role: a closer look will reveal lack of significant clusters during July and August for both 2009 and 2010. This annual pattern can be explained by the practice of taking summer holidays during these months. The combination of the space-time LISA, which results in clusters significant both over time and space, and the visual layout of the Space-Time Calendar enables us to pick up this kind of patterns in an intuitive way.

Interesting as they may have been, the above two examples confirmed patterns that a careful observer of Amsterdam may have also noticed and therefore speak to the validity of the Space-Time Calendars. The discussion about the next two neighbourhoods will exemplify how our visualisation device can reveal changes in the city almost as they occur. Figure 5 (c) presents the Space-Time Calendar for Bos en Lommerplein in the west, which is the local centre of a broader multicultural area (Bos en Lommer). This neighbourhood performs central functions throughout the whole study period as it can be seen in Figure 5. However, the frequency of the occurrence of such occurrences increases over time. While in the beginning of our study period the area was a hotspot sporadically, in 2010 Bos en Lommerplein started performing central functions almost every day. The Space-Time Calendar makes visually explicit the transformation of Bos en Lommerplein and its increasing role as a centre of human activity. These increase was due to a regeneration plan brought in effect around 2008/10. As documented by Broitman and Koomen (2015), the area experienced a clear residential intensificationduring the period 2000-2010, and a newly developed commercial area came into effect towards the end of the decade.

A slightly different case can be observed in Andreasplein. This neighbourhood is located in the south-west part of our study area and it is placed between the south end of Vondelpark and Rembrandtpark and is adjacent to the A10 motorway. Its location between areas whith the potential to attract human activity becomes apparent in the left end of Figure 5 (d), where the most commonly observed cluster in the Space-Time Calendar is the low-high, depicted in light blue. This reflects GSM zones with low density of mobile phone usage which are adjacent to GSM zones with high values of density. In other



(a) Leidseplein

(b) Zuidas

(c) Bos en Lommerplein

(d) Andreaseplein

Colour scheme: dark red for High-High significant clusters; dark blue for Low-Low significant clusters; light red for High-Low significant spatial outliers; light blue for Low-High significant spatial outliers; dark grey for non-significant.

Figure 5: Space-Time Calendars

words, Andreasplein is an area with low activity but nearby some hotspots. This situation changes within the period of analysis. Progressively, but specially from the end of 2009 onwards, the area is no longer low activity near high, but it becomes a hotspot in itself. This process can be seen in the Space-Time Calendar as the amount of red increases substantially as one looks right. This signifies the evolution of Andreasplein as it morphs from an in-between location to a cluster in its own right over the course of almost three years. This phenomenon is essentially the development of a spatial spillover of activity, and it is captured with high resolution and made explicit by the Space-Time Calendar.

Conclusions

Driven by the availability of new forms of data, this paper pairs such data with a space-time methodological framework to enhance our understanding of the dynamic structure of cities. It compares the ability of two different families of clustering methods and their space-time extensions, LISA and STSS, in understanding the fast and slow dynamics of cities. More importantly, the paper proposes the Space-Time Calendar, a novel approach which enables to overcome some of the challenges researchers face when trying to understand the dynamics of cities. Urban changes happen at various temporal scales in an interconnected way, but researchers tend to study these scales in isolation. The Space-Time Calendar removes some of these barriers and enables the study of different temporal scales of urban dynamics in an integrated manner. In essence, the Space-Time Calendar provides a platform to study the fast dynamics of cities as they change slowly over longer periods of time. We illustrate these aspects with an empirical application using mobile phone data for the city of Amsterdam. Given the inherent methodological difficulty in considering an explicit space-time perspective, our approach offers a direct contribution to research fields whose mission is to understand cities and their dynamics. Our methodological proposal directly contributes to urban science and geographic data science, particularly to the space-time analytics field, which understands and analyses cities as systems, the functional properties of which evolve over time.

The paper delivers two key contributions. First, we illustrate a clear advantage in closely connecting the choice of methods to the nature of the data, both over space and time. In our particular example, this implied aerial data with a clear temporal dimension. As such, the best match proved to be the Space-Time LISA: an approach that explicitly includes both space and time into the analysis, formalizing the spatial dimension of the data as a discrete partition of the geography. The main advantages of this approach are the afforded flexibility in defining space with a spatial weights matrix, as well as being able to accommodate the daily rhythm of the city, which we would not be able to do with any of the other techniques considered.

The application of our methodological proposal and the introduction of the Space-Time Calendar leads to the second and main contribution of the paper: an enhanced view into the dynamics of cities and how neighbourhoods and their functions evolve over time thanks to the Space-Time Calendar. Traditionally, empirical urban research has been more successful in coping with slower changes than with the faster dynamics of cities. The recent abundance of new sources of data has not been supported by the development of methods capable of utilising their full potential. Our approach finds clusters which are statistically significant not only over space, but also across time. Moreover, the spatial and temporal granularity of the data in combination with the explicit space-time nature of our methodological proposal allows the identification of these activity hubs in a very precise manner. Most importantly, our approach enables the urban researcher to study the evolution of these dynamics over longer periods of time and observe how neighbourhoods and their functions have been transformed in the long run. In other words, the solution we propose arranges information from an arbitrary large number of (space-time) LISA maps in a way that allows to characterize the daily pattern of activity of a given area and, at the same time, to easily spot changes in these patterns over longer periods of time.

Although this paper includes a specific application for the sake of illustration, our contribution goes beyond the particular case of Amsterdam and mobile phone data: both the methodological framework and, specially, the Space-Time Calendar are easily applicable to other cities and can be used in combination with alternative sources of data. Hence, our approach speaks directly to the concerns raised by urban scholars in regards to the lack of a dynamic understanding of the structure of cities. In doing so, we directly engage with the methodological gaps that new forms of data pose for traditional quantitative urban research. As such, this represents a step forward in Kitchin's call for work to "adapt, hone and extend existing techniques and to invent new methods that can make sense of and extract value from big data and data infrastructures" (Kitchin, 2014b, Ch. 6). From a broader perspective, our contribution also advances what Shelton et al. (2014) termed as data-driven urban governance. In effect, using new forms of data, coupled with appropriate methods, to effectively monitor cities can empower urban planning. For example, the methodology proposed in this paper could be applied to streamed data, enabling local authorities to monitor such changes relevant to their core mission in a near real-time manner. It is the intersection of these two worlds, urban data and policy, where the application of methods like the one proposed in this paper can have the most promising impact.

References

a Yan, N., Mei, C.-L., and Wang, N. (2015). A unified bootstrap test for local patterns of spatiotemporal association. *Environment and Planning A*, 47(1):227–242.

Alonso, W. (1964). Location and land use. Toward a general theory of land rent. *Location and land use. Toward a general theory of land rent.*

An, L., Tsou, M.-H., Crook, S. E. S., Chun, Y., Spitzberg, B., Gawron, J. M., and Gupta, D. K. (2015). Space–time analysis: Concepts, quantitative methods, and future directions. *Annals of the Association of American Geographers*, 105(5):891–914.

Anas, A., Arnott, R., and Small, K. A. (1998). Urban spatial structure. *Journal of economic literature*, pages 1426–1464.

Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical analysis*, 27(2):93–115.

Anselin, L., Syabri, I., and Kho, Y. (2006). Geoda: an introduction to spatial data analysis. *Geographical analysis*, 38(1):5–22.

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49:45–53.

Arribas-Bel, D. (2015). The spoken postcodes. *Regional Studies, Regional Science*, 2(1):458–461.

Arribas-Bel, D., Kourtit, K., Nijkamp, P., and Steenbruggen, J. (2015). Cyber cities: social media as a tool for understanding cities. *Applied Spatial Analysis and Policy*, 8(3):231–247.

Arribas-Bel, D. and Sanz-Gracia, F. (2014). The validity of the monocentric city model in a polycentric age: US metropolitan areas in 1990, 2000 and 2010. *Urban Geography*, 35(7):980–997.

Arribas-Bel, D. and Schmidt, C. R. (2013). Self-organizing maps and the US urban spatial structure. *Environment and Planning B: Planning and Design*, 40(2):362–371.

Batty, M. (2009). Cities as complex systems: Scaling, interaction, networks, dynamics and urban morphologies. In *Encyclopedia of complexity and systems science*, pages 1041–1071. Springer.

Batty, M. (2013a). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3):274–279.

Batty, M. (2013b). Visually-Driven Urban Simulation: exploring fast and slow change in residential location. *Environment and Planning A*, 45(3):532–552.

Batty, M. et al. (2012). Smart cities, big data. *Environment and Planning-Part B*, 39(2):191.

Baumont, C., Ertur, C., and Gallo, J. (2004). Spatial analysis of employment and population density: the case of the agglomeration of Dijon 1999. *Geographical Analysis*, 36(2):146–176.

Bettencourt, L. M. (2014). The uses of big data in cities. *Big Data*, 2(1):12–22.

Blondel, V. D., Decuyper, A., and Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(10):1–55.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebre, E. (2008). Fast unfolding of community hierarchies in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10:P1000.

Boeing, G. and Waddell, P. (2016). New insights into rental housing markets across the united states: Web scraping and analyzing craigslist rental listings. *Journal of Planning Education and Research*.

boyd, D. and Crawford, K. (2012). Critical questions for big data. information. *Communication & Society*, 15(5):662–679.

Broitman, D. and Koomen, E. (2015). Residential density change: Densification and urban expansion. *Computers, Environment and Urban Systems*, 54:32–46.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J., and Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26:301–313.

Calabrese, F., Ferrari, L., and Blondel, V. D. (2014). Urban Sensing Using Mobile Phone Network Data: A Survey of Research. *ACM Computing Surveys (CSUR)*, 47(2):25.

Cheng, T. and Adepeju, M. (2014). Modifiable temporal unit problem (mtup) and its effect on space-time cluster detection. *PloS one*, 9(6):e100465.

Cheng, T. and Wicks, T. (2014). Event detection using twitter: a spatio-temporal approach. *PloS one*, 9(6):e97807.

Christaller, W. (1966). *Central places in southern Germany*. Prentice-Hall.

Delmelle, E., Dony, C., Casas, I., Jia, M., and Tang, W. (2014). Visualizing the impact of space-time uncertainties on dengue fever patterns. *International Journal of Geographical Information Science*, 28(5):1107–1127.

Dodge, M. and Kitchin, R. (2005). Codes of life: Identification codes and the machine-readable world. *Environment and Planning D: Society and Space*, 23(6):851–881.

Ferrari, L., Rosi, A., Mamei, M., and Zambonelli, F. (2011). Extracting urban patterns from location-based social networks. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pages 9–16. ACM.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221.

Gordon, P. and Richardson, H. W. (1996). Beyond polycentricity: the dispersed metropolis, Los Angeles, 1970-1990. *Journal of the American Planning Association*, 62(3):289–295.

Guillain, R., Le Gallo, J., and Boiteux-Orain, C. (2006). Changes in spatial and sectoral patterns of employment in Ile-de-France, 1978-97. *Urban Studies*, 43(11):2075–2098.

Hagerstrand, T. et al. (1968). Innovation diffusion as a spatial process. *Innovation diffusion as a spatial process*.

Han, S. S. (2005). Polycentric urban development and spatial clustering of condominium property values: Singapore in the 1990s. *Environment and Planning A*, 37(3):463–481.

Jacobs-Crisioni, C., Rietveld, P., Koomen, E., and Tranos, E. (2014). Evaluating the impact of land-use density and mix on spatiotemporal urban activity patterns: An exploratory study using mobile phone data. *Environment and Planning A*, 46(11).

Kang, H. (2010). Detecting agglomeration processes using space–time clustering analyses. *The Annals of Regional Science*, 45(2):291–311.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3):262–267.

Kitchin, R. (2014a). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1):2053951714528481.

Kitchin, R. (2014b). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics-Theory and methods*, 26(6):1481–1496.

Kulldorff, M., Heffernan, R., Hartman, J., Assunçao, R., and Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. *Plos med*, 2(3):e59.

Kulldorff, M., Huang, L., and Konty, K. (2009). A scan statistic for continuous data based on the normal probability model. *International journal of health geographics*, 8(1):1.

Kwan, M.-P. (2013). Beyond space (as we knew it): toward temporally integrated geographies of segregation, health, and accessibility: Space–time integration in geography and GIScience. *Annals of the Association of American Geographers*, 103(5):1078–1086.

Laney, D. (2001). 3d data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6:70.

Lee, C., Devillers, R., and Hoeber, O. (2014). Navigating spatio-temporal data with temporal zoom and pan in a multi-touch environment. *International Journal of Geographical Information Science*, 28(5):1128–1148.

Lee, J. and Li, S. (2016). Extending moran's index for measuring spatiotemporal clustering of geographic events. *Geographical Analysis*, pages n/a–n/a.

Licoppe, C., Diminescu, D., Smoreda, Z., and Ziemlicki, C. (2008). Using mobile phone geolocalisation for 'socio-geographical' analysis of co-ordination, urban mobilities, and social integration patterns. *Tijdschrift voor Economische en Sociale Geografie*, 99(5):584–601.

López, F. A., Chasco, C., and Gallo, J. L. (2015). Exploring scan methods to test spatial structure with an application to housing prices in madrid. *Papers in Regional Science*, 94(2):317–346.

Louail, T., Lenormand, M., Cantu Ros, O. G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. J., and Barthelemy, M. (2014). From mobile phone data to the spatial structure of cities. *Scientific reports*, 4:5276.

Marz, N. and Warren, J. (2015). *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications Co.

Mayer-Schonberger, V. and Cukier, K. (2013). Big data: A revolution that will change how we live, work and think. *London: John Murray. ISBN*, 978:0544227750.

McDonald, J. F. (1987). The identification of urban employment subcenters. *Journal of Urban Economics*, 21(2):242–258.

McMillen, D. P. (2001). Nonparametric employment subcenter identification. *Journal of Urban Economics*, 50(3):448–473.

Mills, E. S. (1972). Studies in the Structure of the Urban Economy.

Muth, R. F. (1969). Cities and housing; the spatial pattern of urban residential land use.

Páez, A. and Scott, D. M. (2005). Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, 61(1):53–67.

Paez, A., Uchida, T., and Miyamoto, K. (2001). Spatial association and heterogeneity issues in land price models. *Urban Studies*, 38(9):1493–1508.

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S.-L., Li, T., and Zhou, C. (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 28(9):1988–2007.

Pereira, R. H. M., Nadalin, V., Monasterio, L., and Albuquerque, P. H. M. (2013). Urban centrality: a simple index. *Geographical Analysis*, 45(1):77–89.

Portugali, J. (2011). *Complexity, cognition and the city*. Springer-Verlag, Berlin.

Rabari, C. and Storper, M. (2014). The digital skin of cities: urban theory and research in the age of the sensored and metered city, ubiquitous computing and big data. *Cambridge Journal of Regions, Economy and Society*, page rsu021.

Rae, A. (2015). Online housing search and the geography of submarkets. *Housing Studies*, 30(3):453–472.

Rae, A. and Sener, E. (2016). How website users segment a city: The geography of housing search in london. *Cities*, 52:140 – 147.

Ratti, C., Pulselli, R. M., Williams, S., and Frenchman, D. (2006). Mobile Landscapes: using location data from cell phones for urban analysis. *Environment & Planning B*, 33(5):727–748.

Reades, J., Calabrese, F., and Ratti, C. (2009). Eigenplaces: Analyzing cities using the space-time structure of the mobile phone network. *Environment and Planning B*, 36(5):824–836.

Reades, J., Calabrese, F., Sevtsuk, A., and Ratti, C. (2007). Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 6(3):30–38.

Rey, S. J. and Janikas, M. V. (2006). Stars: space–time analysis of regional systems. *Geographical analysis*, 38(1):67–86.

Richardson, D. B. (2013). Real-time space–time integration in GIScience and geography: Space–time integration in geography and GIScience. *Annals of the Association of American Geographers*, 103(5):1062–1071.

Roth, C., Kang, S. M., Batty, M., and Barthélemy, M. (2011). Structure of urban movements: polycentric activity and entangled hierarchical flows. *PloS one*, 6(1):e15923.

Salvati, L. and Carlucci, M. (2014). Distance matters: Land consumption and the mono-centric model in two southern European cities. *Landscape and Urban Planning*, 127:41–51.

Salvati, L., Venanzoni, G., Serra, P., and Carlucci, M. (2016). Scattered or polycentric? Untangling urban growth in three southern European metropolitan regions through exploratory spatial data analysis. *The Annals of Regional Science*, pages 1–29.

Sevtsuk, A. and Ratti, C. (2010). Does Urban Mobility Have a Daily Routine? Learning from the Aggregate Data of Mobile Networks. *Journal of Urban Technology*, 17(1):41–60.

Shelton, T., Zook, M., and Wiig, A. (2014). The 'actually existing smart city'. *Cambridge Journal of Regions, Economy and Society*, page rsu026.

Snickars, F., Johansson, B., and Leonardi, G. (1982). Nested Dynamics of Metropolitan Processes ad Policies. *Laxenburg: International Institute for Applied Systems Analysis.*

Snow, C. P. (1964). *The two cultures and a second look.* Cambridge University Press, Cambridge.

Steenbruggen, J., Tranos, E., and Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities? *Telecommunications Policy*, 39(3-4).

Steenbruggen, J. G. M., Tranos, E., and Rietveld, P. (2016). Can Motorway Traffic Incidents be detected by Mobile Phone Data? An empirical application in the Netherlands. *Journal of Transport Geography*, 54:81–90.

Toole, J. L., Ulm, M., Gonzalez, M. C., and Bauer, D. (2012). Inferring land use from mobile phone activity. In *UrbComp '12*, pages 1–8, Beijing, China.

Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia.* WW Norton & Company.

Tranos, E. and Nijkamp, P. (2015). Mobile phone usage in complex urban systems: a space-time, aggregated human activity study. *Journal of Geographical Systems*, 17(2).

Tuia, D., Kaiser, C., Da Cunha, A., and Kanevski, M. (2009). Detection of urban socio-economic patterns using clustering techniques. In *Geocomputation and Urban Planning*, pages 19–36. Springer.

Von Thunen, J. H. (1826). Der isolierte Staat in Beziehung auf landtschaft under nationalokonomie. *Hamburg.(English translation by CM Wartenberg. Von Thunen's Isolated State. Oxford: Pergamon Press, 1966).*

Wegener, M., Gnad, F., and Vannahme, M. (1983). The time scale of urban change. In Hutchinson, B. and Batty, M., editors, *Advances in Urban Systems Modelling*, pages 175–197. North-Holland, Amsterdam.

Ye, X. and Carroll, M. C. (2011). Exploratory space-time analysis of local economic development. *Applied Geography*, 31(3):1049–1058.

Zikopoulos, P. C., Eaton, C., DeRoos, D., Deutsch, T., and Lapis, G. (2012). Understanding big data. *New York et al: McGraw-Hill*.